Evaluating strategies to contrast misogynistic online content using Delphi-based scenarios

Simone Di Zio, Lara Fontanella and Alice Tontodimamma

University «G. d'Annunzio» of Chieti-Pescara

ASA Conference

Bologna, 6 September-8 September, 2023

Associazione per la Statistica Applicata

Applied Statistics Association

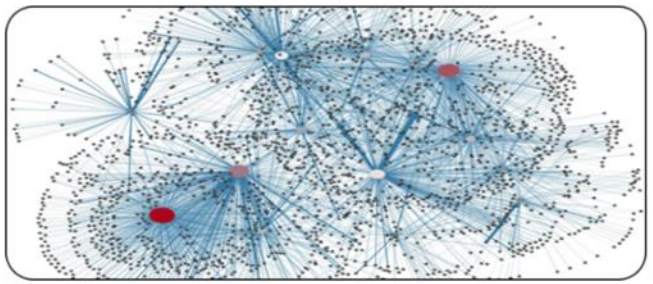
Outline

1. **ICOMIC** project
2. The **aim**
3. **Delphi**-based scenarios approach
4. Selection of How the **experts**
5. Delphi **questionnaire and survey**
6. **Results** of the Delphi
7. **Robust ranking** of Delphi results and **Fuzzy c-means clustering** of the ranked projection
8. Future **Developments**

ICOMIC: Identifying and Countering Online Misogyny

Project funded by EU Next Generation, MUR-Fondo Promozione e Sviluppo-DM 737/2021

Misogynistic speech detection



Identification of producers of misogynistic content

Countering misogynistic speech

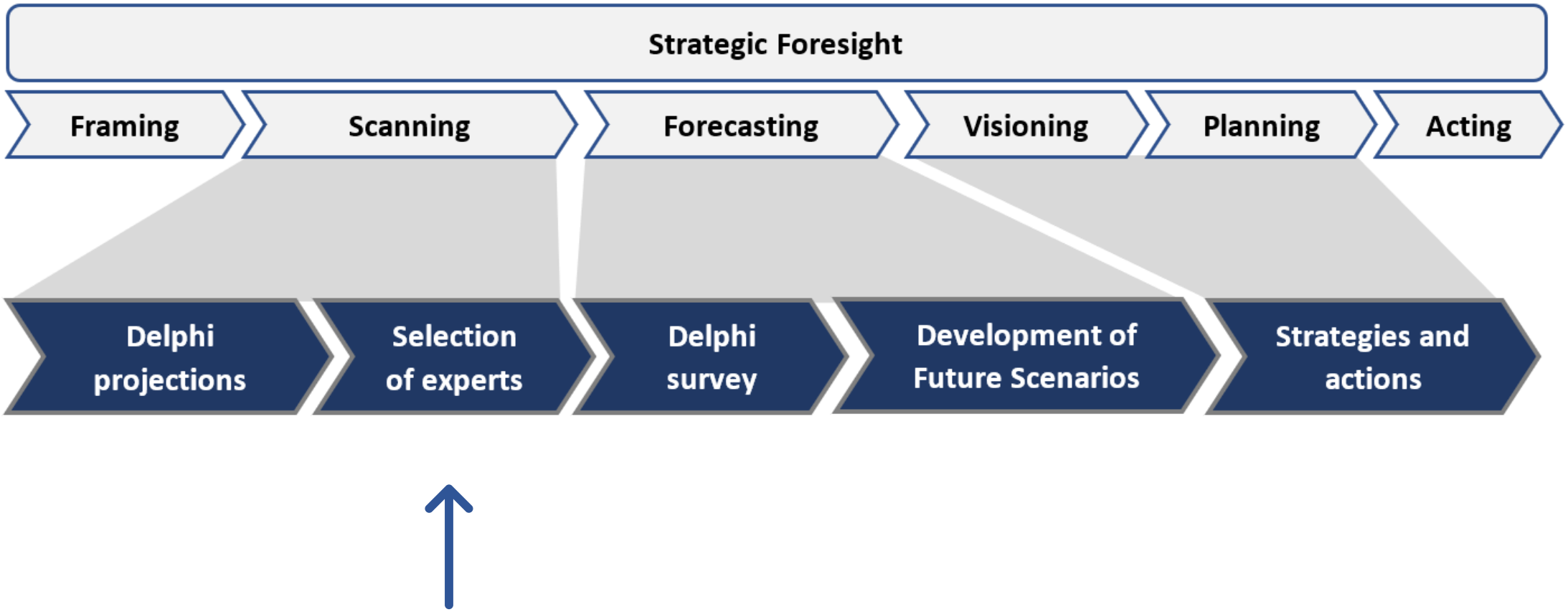
Finanziato dall'Unione europea
NextGenerationEU

https://csrlab.unich.it/en/icomic/

The goal

- In our research, we **aim to investigate long-term strategies** that can be taken **to contain the dissemination of online misogynistic speech**.
- The definition of possible counteractions and their feasibility and effectiveness depend on **how our society evolves in the future**.
- It is useful to generate a number of future scenarios in order to act in the present.
- To address this task, we consider a **Delphi-based Scenarios approach**.

Delphi-based Scenarios approach



How the experts were selected

We followed these steps:

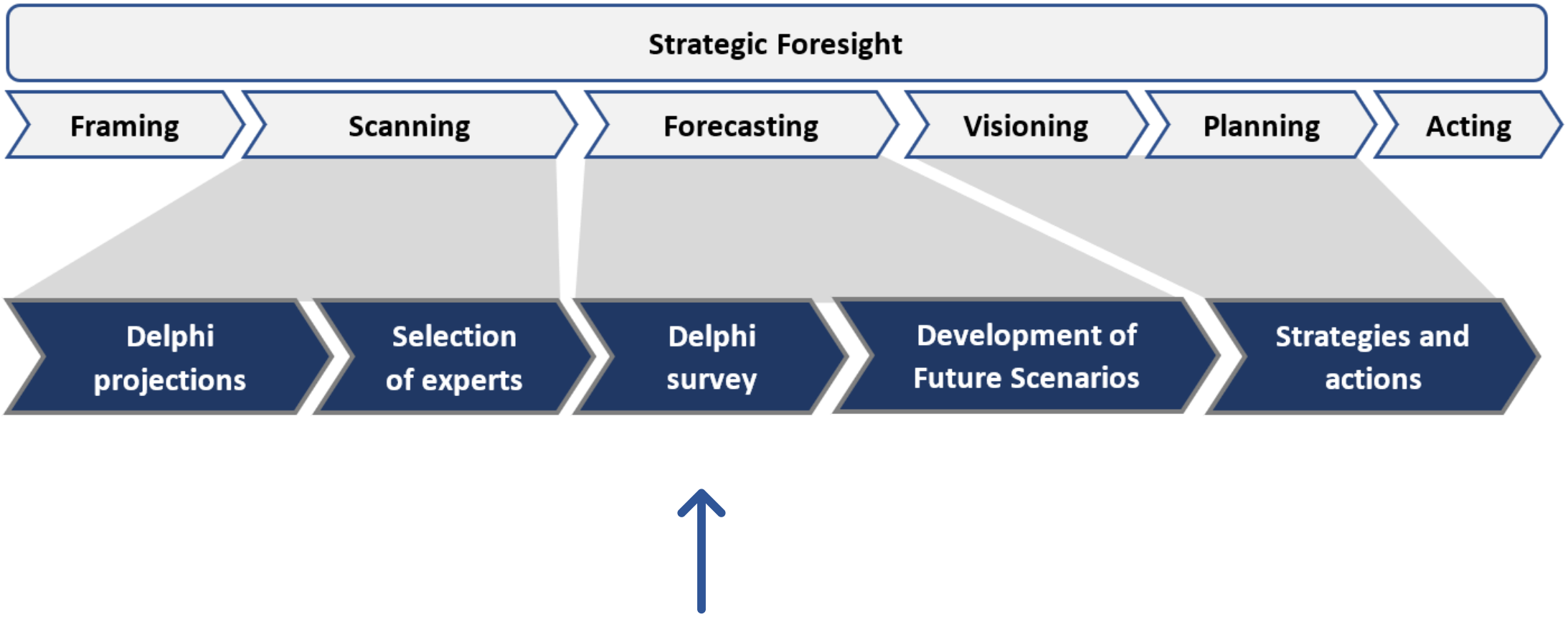
1. **Bibliometric research** (keywords: “misogyn*” and “hate speech”)

2. **Download** the documents that contain these words in the title, abstract or document: 2307 documents on hate speech, 672 authors; 2290 documents on misogyny, 268 authors.

3. For each **author** we calculated **an index**: number of citations / number of publications

4. **1000 experts selected**

Delphi-based Scenarios approach



Delphi questionnaire

- Questionnaire online on Alchemer
- It contains three sections:
  - **FIRST SECTION**:

    21 questions about **countermeasures to date**, evaluated with 3 variables: effectiveness, prevention and probability of existence in 2040;
  - **SECOND SECTION**:

    9 questions about **future counter and preventive actions**, evaluated with 5 variables: probability of occourence in 2040, effectiveness in 2040, impact in 2040, future technological feasibility, future legal feasibility;
  - **THIRD SECTION**:

    19 questions about **key factors of social change**, evaluated with 4 variables: plausibility in 2040, impact in 2040, desirability in 2040, effect in 2040.

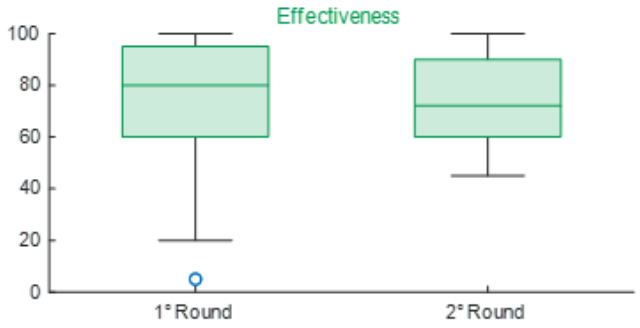
Results of the First and Second Round

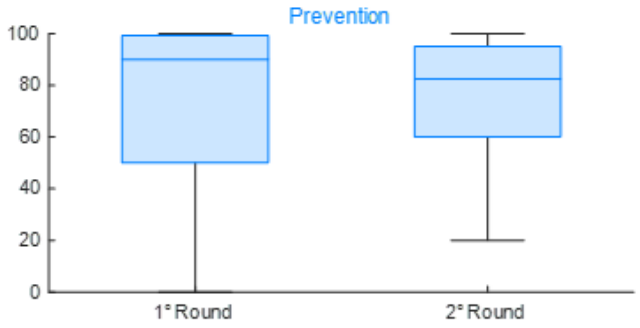
First Round 41 partecipants

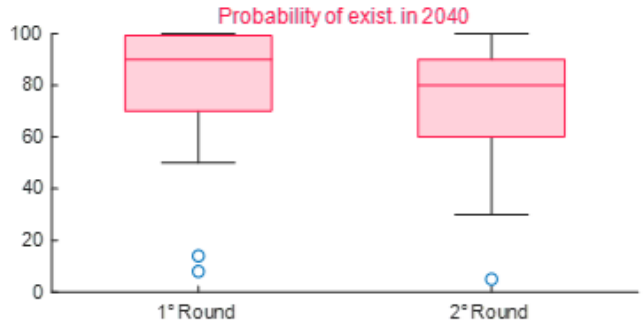
Second Round 30 partecipants

Countermeasures to date: results

2. Educate yourself and others on the harm caused by misogynistic hate speech.

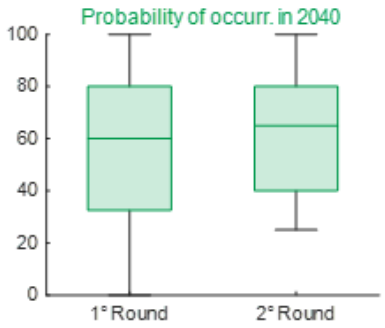


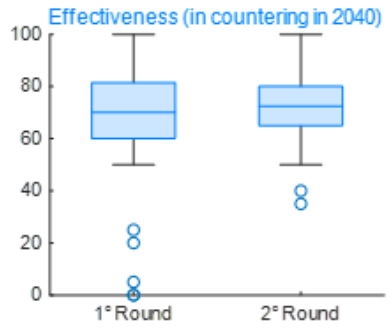


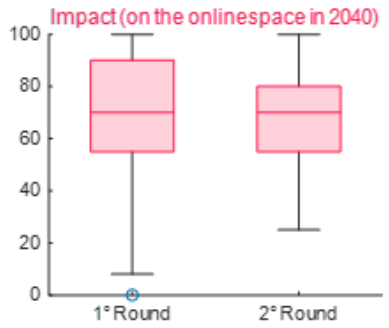


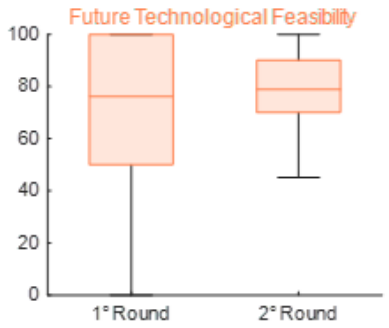
Future counter and preventive actions: results

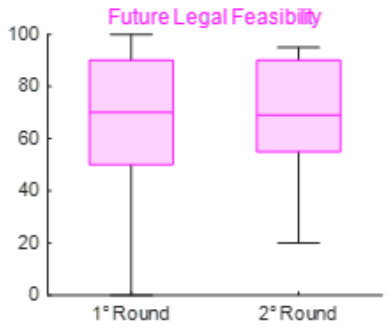
1. Media literacy is a compulsory subject in school.





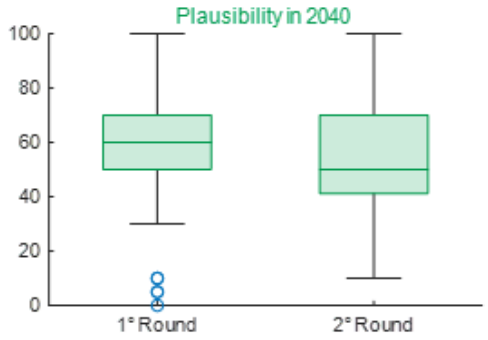


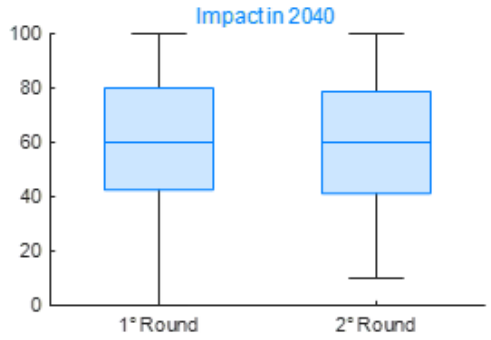


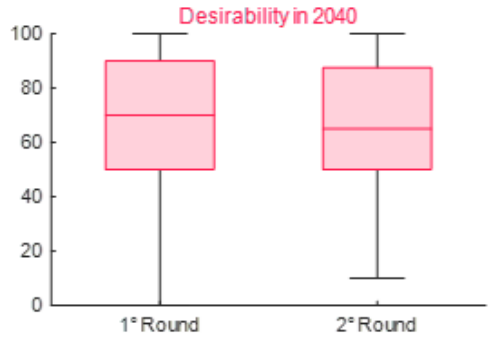


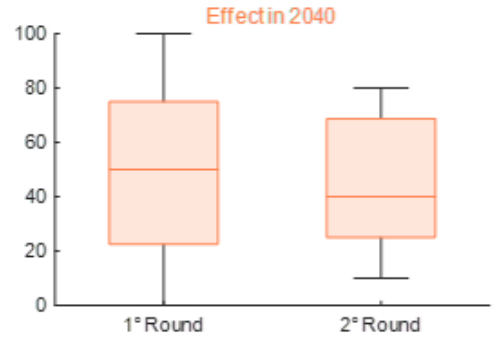
Key factors of social change: results

3. Legislation on freedom of expression will offer new possibilities to block haters.

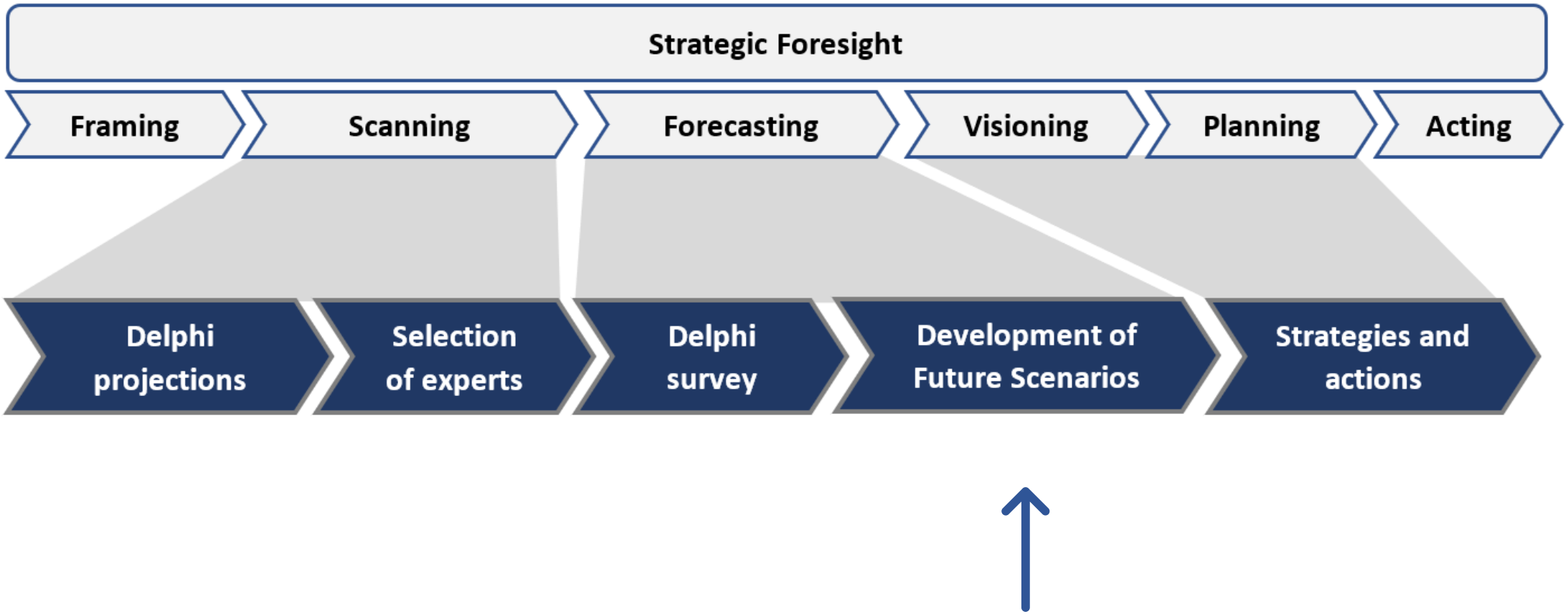








Delphi-based Scenarios approach



Robust Ranking Method

Method to generate ranks of Delphi items independent of Expertise level of Experts (Di Zio et al., 2021).

This method overcomes problems related to:

- **Measurement of skills** (self-assessments and cognitive biases);
- Use of an adequate **weighting system** to calibrate the results at the different levels of expertise;
- **Length of the Delphi questionnaire** (in case of self-rating), since it avoids self-rating questions.
- Improper use of **Euclidean distance**.

Robust Ranking Method

Different normalization and aggregation formulas can be used; based on these choices the ranking changes.

With a **Monte Carlo procedure**, **all formulas can be used simultaneously**, so the results do not depend on the choice of a particular formula.

Furthermore, one can model the **weights $\mathbf{W_e}$** which are in fact estimates.

At each Monte Carlo iteration, we generate a set of weights

$$W_1 + \eta, \dots, W_e + \eta, \dots, W_E + \eta$$

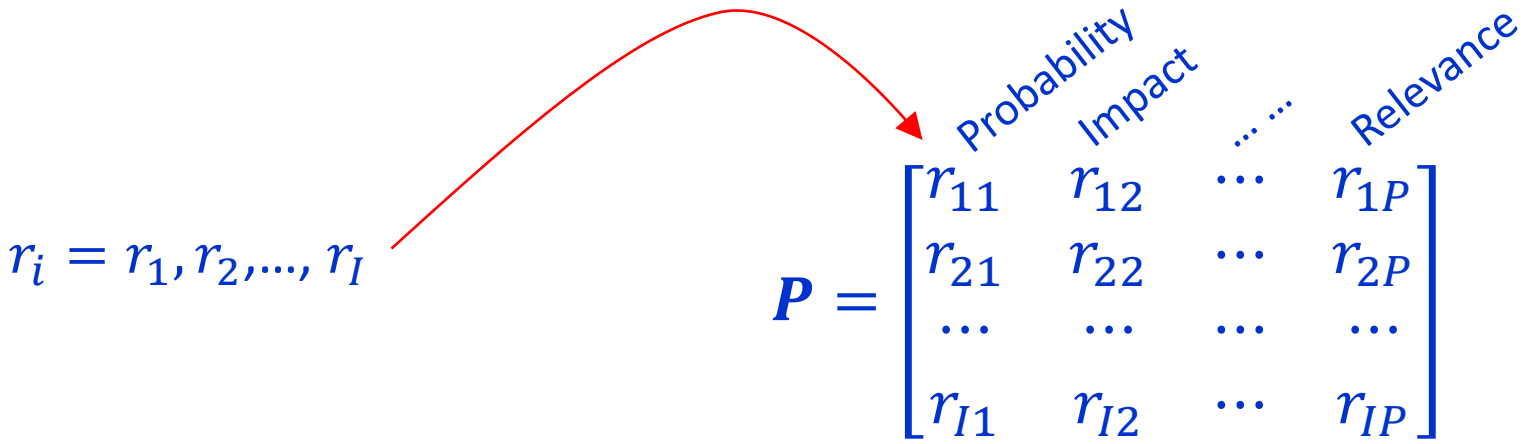
where $\eta \sim N(0, SD_W)$.

Ranks Fuzzy clustering

The final vector of ranks $r_1, r_2, \dots, r_I$ can be used as input into a clustering algorithm.

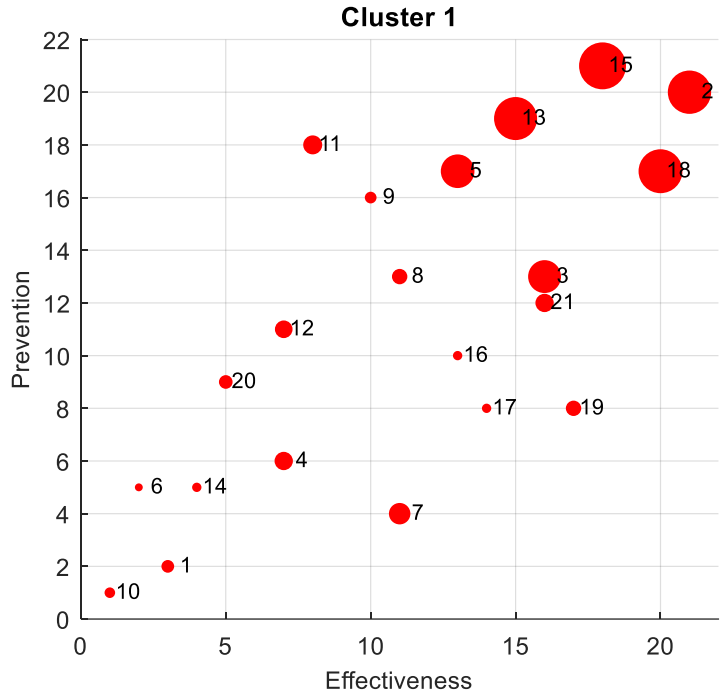
In the case of $P > 1$ variables, the column vector r_{ip} $(p = 1, \dots, P)$ form a usable matrix $\boldsymbol{P}$ $(I \times P)$ for clustering (I: number of items).

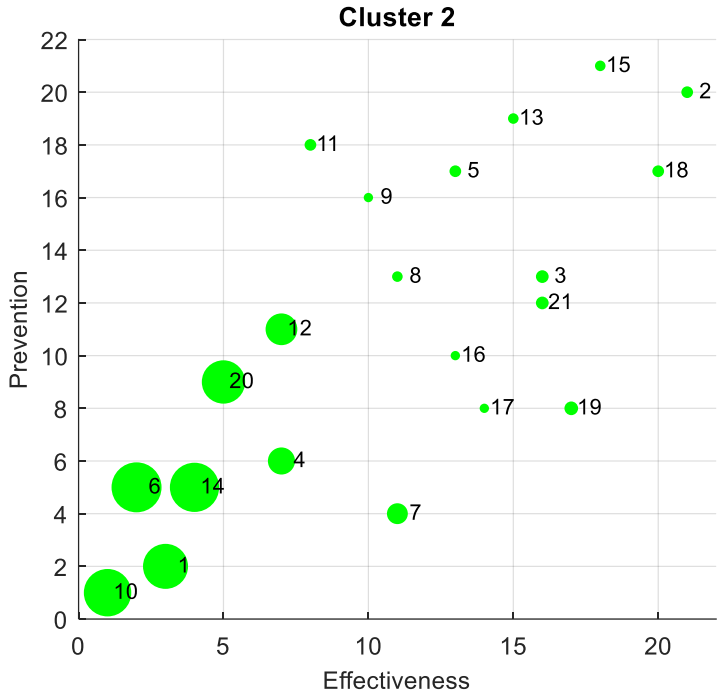
In this method, the fuzzy clustering (c-means) approach is recommended given that there may be items sharing different clusters/scenarios.

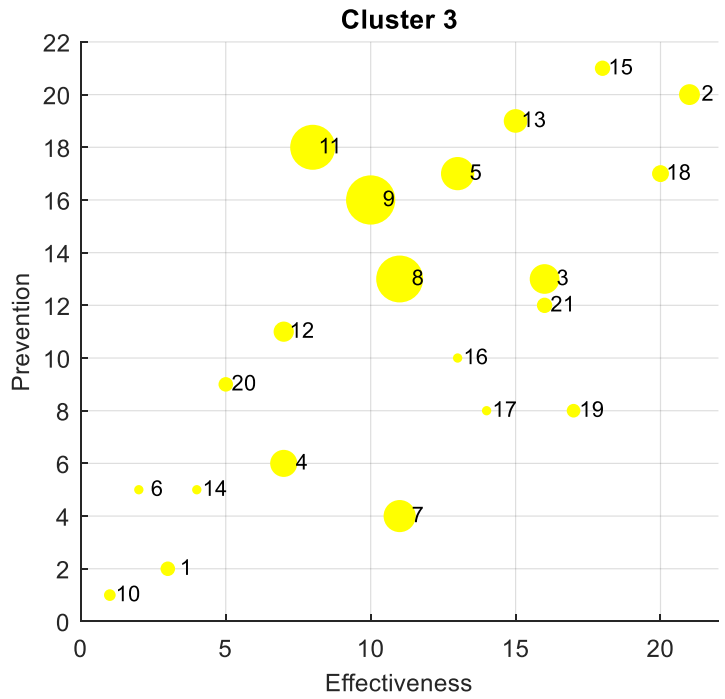


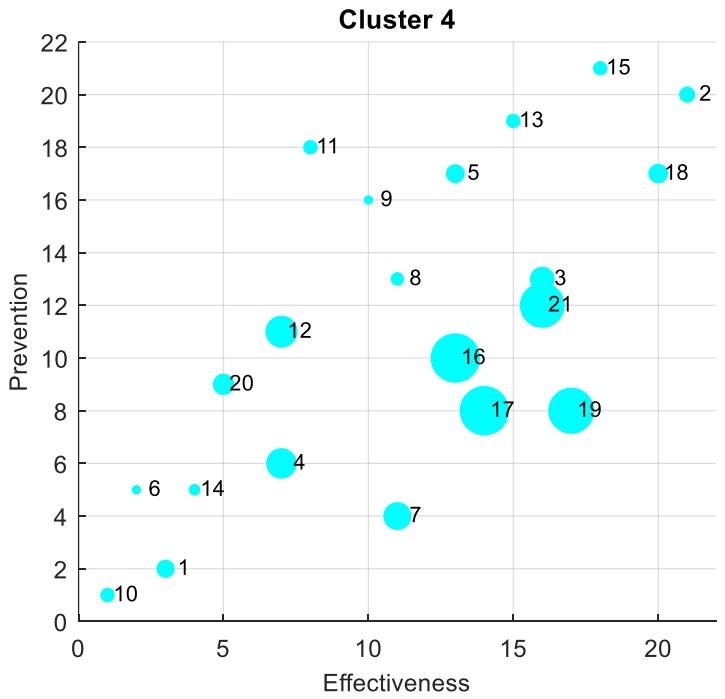
Fuzzy clustering reulsts: **Contermeasures to date**

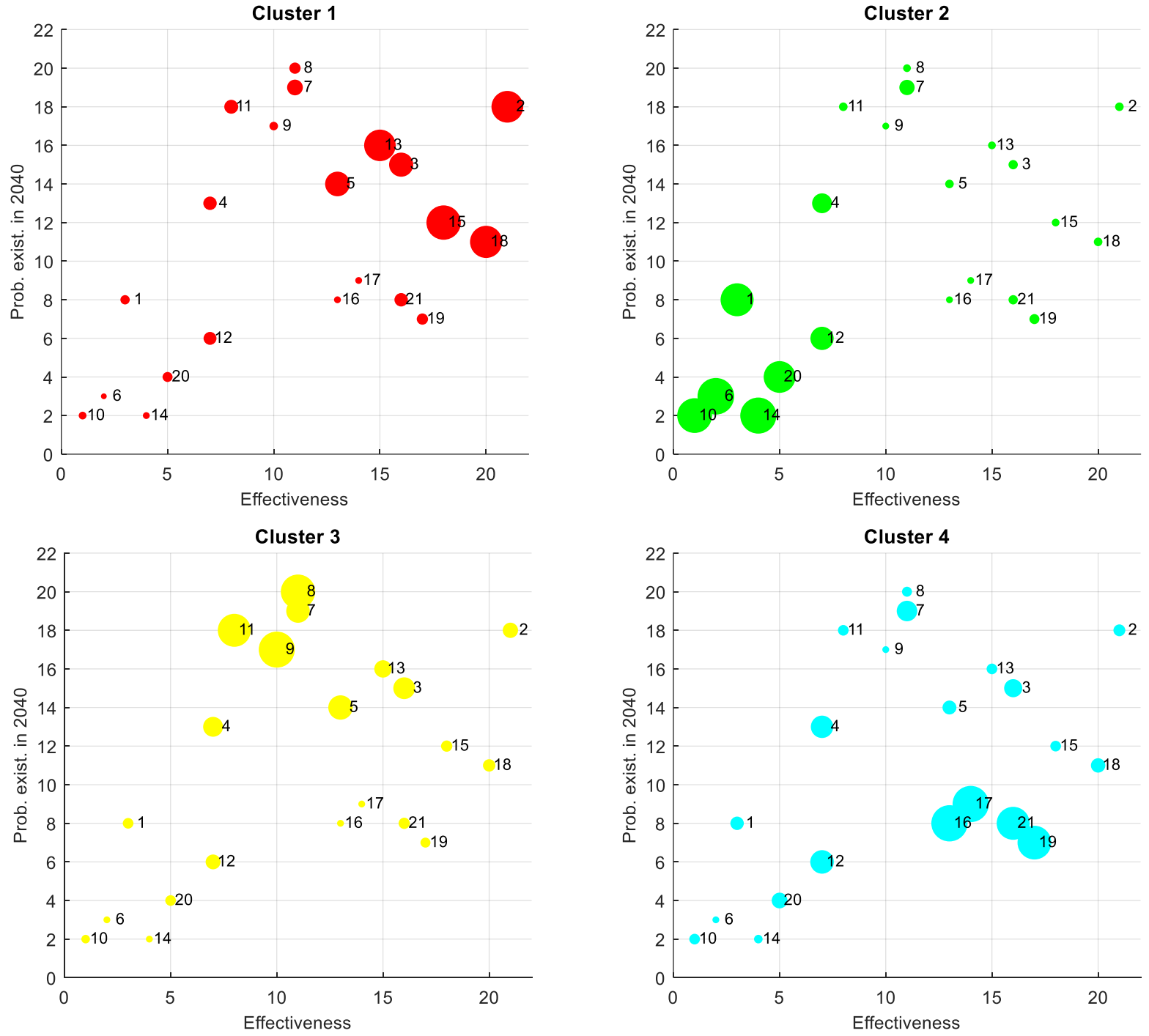
Clus 1	Clus 2	Clus 3	Clus 4	
0.04	0.79	0.06	0.11	1. Report misogynistic hate content to authorities or platform administrators.
0.73	0.03	0.15	0.08	2. Educate yourself and others on the harm caused by misogynistic hate speech.
0.41	0.04	0.33	0.22	3. Speak out against misogynistic hate speech, including publicly condemning it.
0.11	0.27	0.27	0.35	4. Promote and amplify the voices of marginalized communities.
0.43	0.03	0.42	0.12	5. Support and advocate for policies that restrict misogynistic hate speech.
0.00	0.98	0.01	0.01	6. Engage in positive dialogue with those who engage in misogynistic hate speech.
0.16	0.15	0.39	0.29	7. Provide support to those who are targeted by misogynistic hate speech.
0.07	0.02	0.86	0.05	8. Create and share positive content that challenges misogynistic hate speech.
0.03	0.01	0.95	0.01	9. Organize and participate in events and activities that promote diversity and inclusiveness.
0.02	0.88	0.03	0.06	10. Use humor to diffuse and challenge misogynistic hate speech.
0.12	0.03	0.79	0.06	11. Advocate for media literacy and critical thinking skills.
0.10	0.38	0.14	0.39	12. Encourage bystander intervention when misogynistic hate speech is witnessed.
0.72	0.02	0.20	0.06	13. Raise awareness through social media and online activism.
0.01	0.95	0.01	0.03	14. Support victims of misogynistic hate speech through financial or legal means.
0.86	0.02	0.07	0.06	15. Foster community-building activities to promote unity and respect.
0.01	0.01	0.01	0.96	16. Counterspeech (e.g., denouncing speech, Warning of consequences, Humor)
0.01	0.01	0.01	0.96	17. Remove the misogynistic content within a timeframe of 24 h and report on the measures taken.
0.75	0.03	0.09	0.13	18. Legal recourse.
0.07	0.05	0.05	0.83	19. The Social Media Platform removes the hateful material.
0.05	0.73	0.06	0.16	20. The Social Media Platform sends warnings to the hater.
0.11	0.04	0.07	0.79	21. The Social Media Platform restricts/bans the activity of the hater on the platform.


Cluster 1
Prevention
Effectiveness
0 2 4 6 8 10 12 14 16 18 20 22
0 5 10 15 20
15 2 13 11 5 18 9 8 3 21 12 16 20 17 19 4 6 14 7 1 10


Cluster 2
Prevention
Effectiveness
0 2 4 6 8 10 12 14 16 18 20 22
0 5 10 15 20
15 2 13 11 5 18 9 8 3 21 12 16 20 17 19 4 6 14 7 1 10

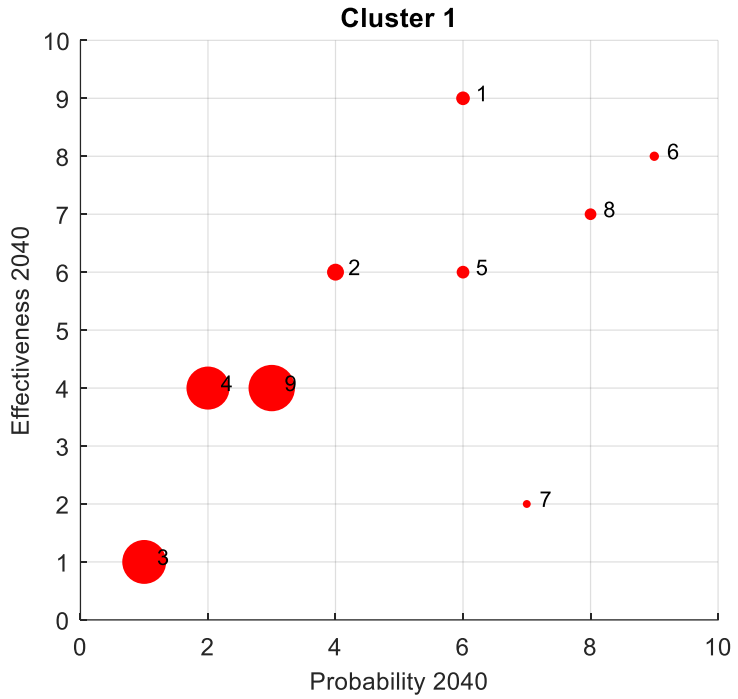

Cluster 3
Prevention
Effectiveness
0 2 4 6 8 10 12 14 16 18 20 22
0 5 10 15 20
15 2 13 11 5 18 9 8 3 21 12 16 20 17 19 4 6 14 7 1 10

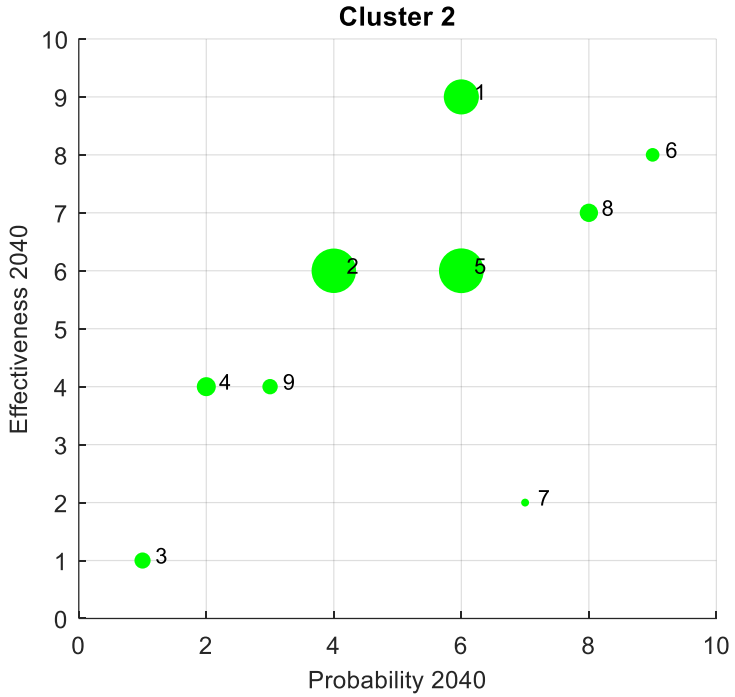

Cluster 4
Prevention
Effectiveness
0 2 4 6 8 10 12 14 16 18 20 22
0 5 10 15 20
15 2 13 11 5 18 9 8 3 21 12 16 20 17 19 4 6 14 7 1 10


Cluster 1
Cluster 2
Cluster 3
Cluster 4
Prob. exist. in 2040
Effectiveness
0
2
4
6
8
10
12
14
16
18
20
22
0
5
10
15
20

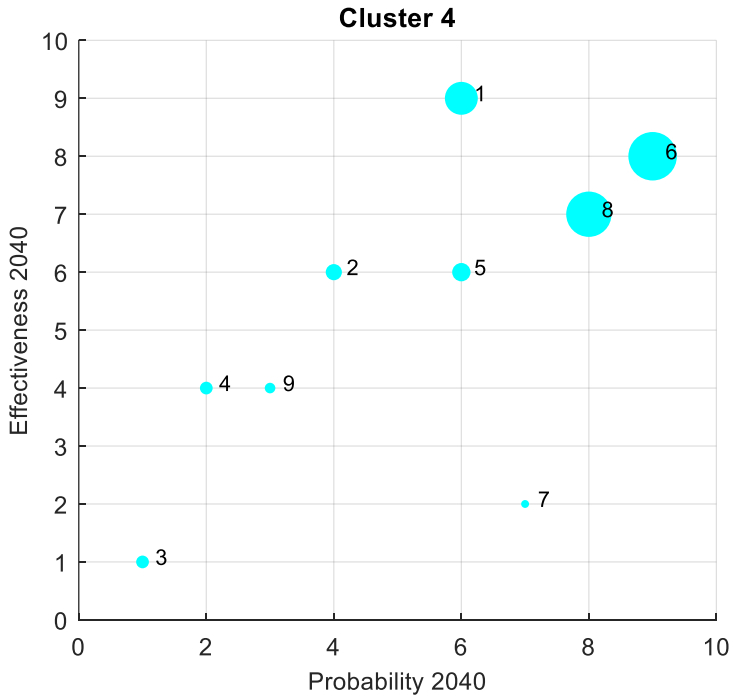
Fuzzy clustering reulsts: **Future counter and preventive actions**

Clus 1	Clus 2	Clus 3	Clus 4	
0.05	0.47	0.06	0.41	1. Media literacy is a compulsory subject in school.
0.09	0.77	0.06	0.08	2. Haters must pay damages to the victims of their posts.
0.75	0.08	0.12	0.04	3. People pay to leave comments online.
0.72	0.12	0.12	0.04	4. Social media platforms with more than 1ml users has transparent proced. to moderate misog. h.s.
0.04	0.78	0.06	0.11	5. Education initiat. (media and inf. literacy program.) are consid. extra-legal counter. and focus on long-term prevent.
0.01	0.05	0.02	0.92	6. Artificial Intelligence handles misogynistic h.s. and warns in real-time the authorities to take countermeasures.
0.00	0.00	1.00	0.00	7. Artificial Intelligence produces misogynistic hate speech without human intervention.
0.03	0.11	0.07	0.80	8. Artificial Intelligence identifies misogynistic hate speech and automatically produces the best counteraction.
0.84	0.07	0.06	0.02	9. Arts education in school is used to countering misogynistic hate speech.

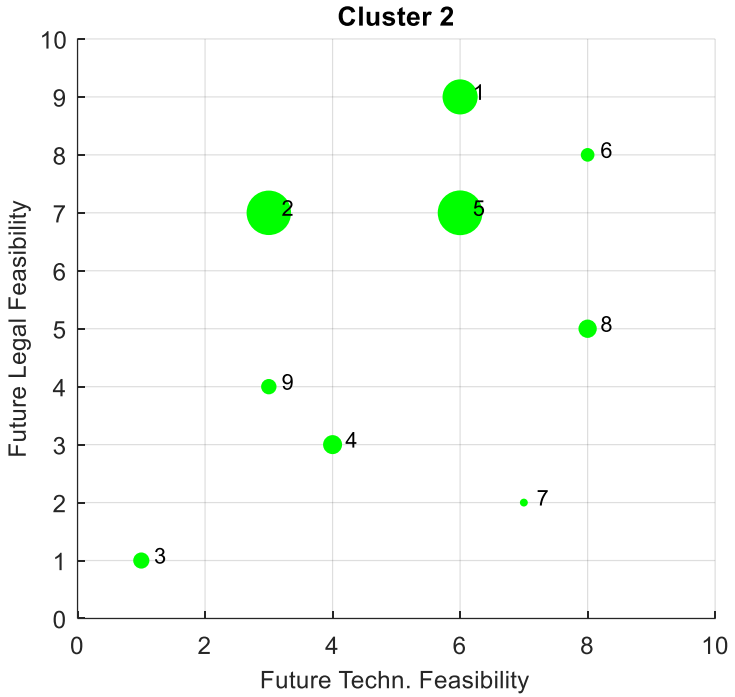

Cluster 1
Effectiveness 2040
Probability 2040
0 1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9


Cluster 2
Effectiveness 2040
Probability 2040
0 1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9

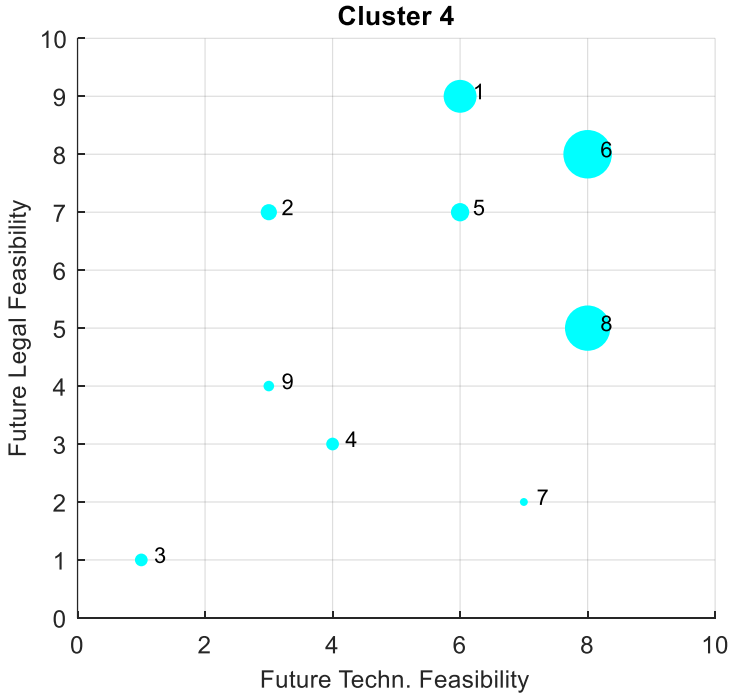
Cluster 3
Effectiveness 2040
Probability 2040
0 1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9

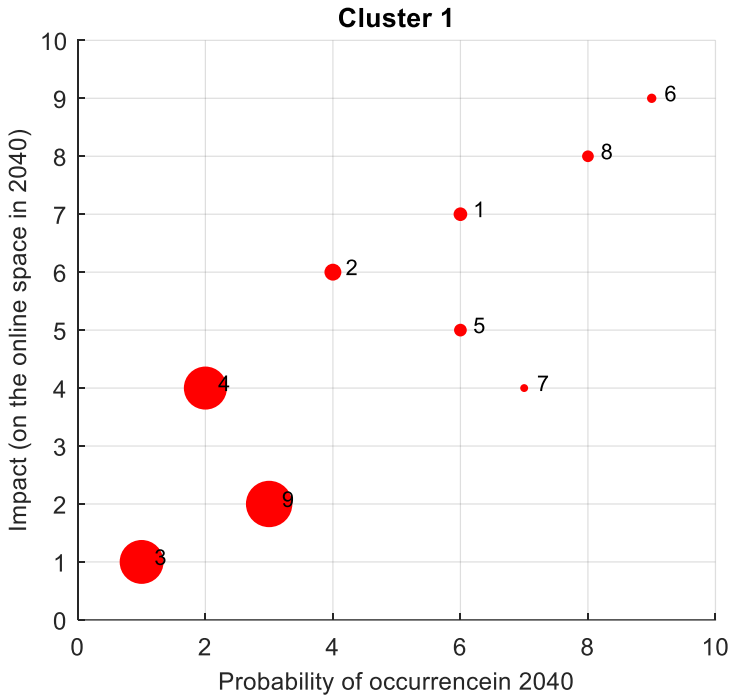

Cluster 4
Effectiveness 2040
Probability 2040
0 1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9

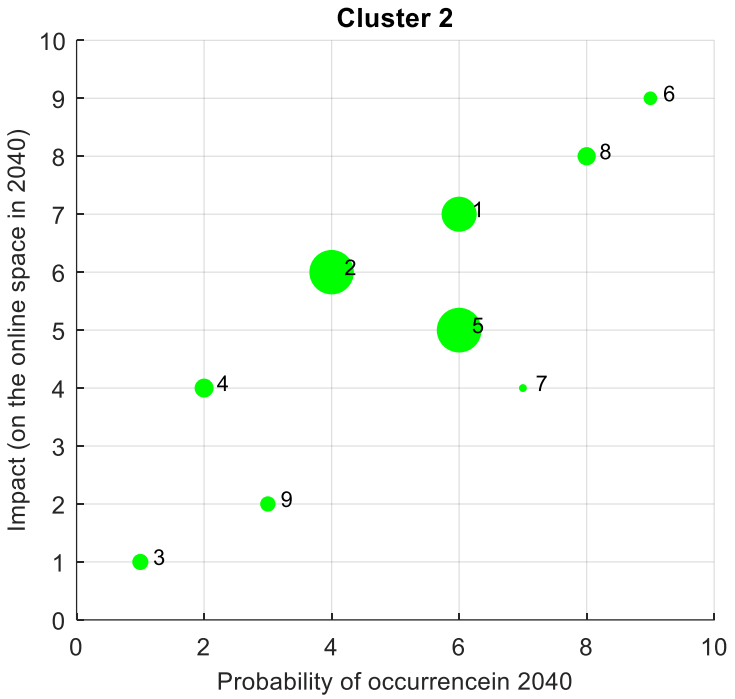
Cluster 1
Future Legal Feasibility
Future Techn. Feasibility
0
1
2
3
4
5
6
7
8
9
10

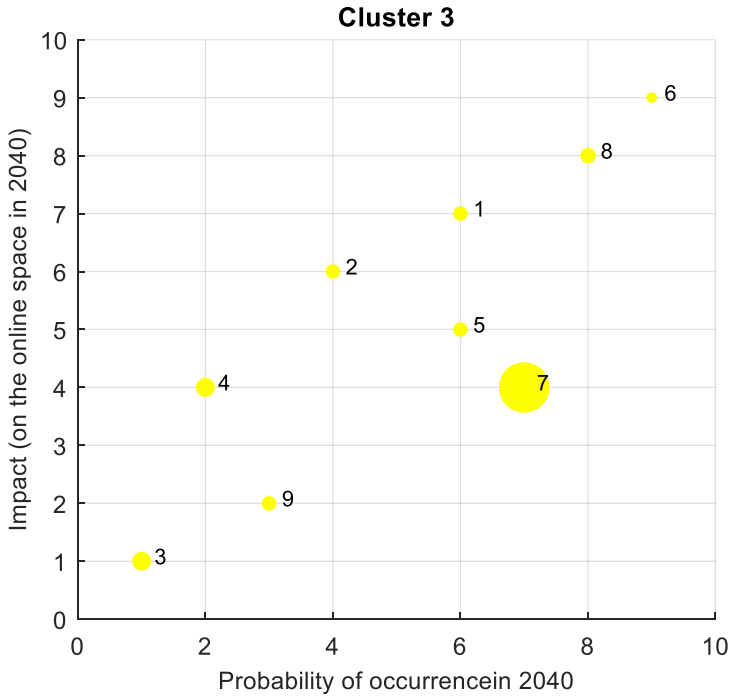

Cluster 2
Future Legal Feasibility
Future Techn. Feasibility
0
1
2
3
4
5
6
7
8
9
10

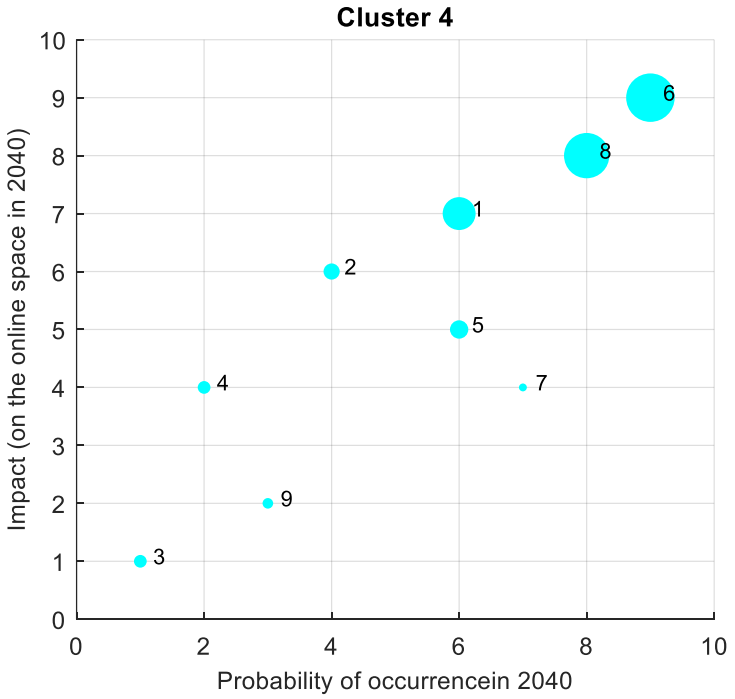
Cluster 3
Future Legal Feasibility
Future Techn. Feasibility
0
1
2
3
4
5
6
7
8
9
10


Cluster 4
Future Legal Feasibility
Future Techn. Feasibility
0
1
2
3
4
5
6
7
8
9
10


Cluster 1
Impact (on the online space in 2040)
Probability of occurrencein 2040
0 1 2 3 4 5 6 7 8 9 10
6
8
1
2
5
4
7
9
3

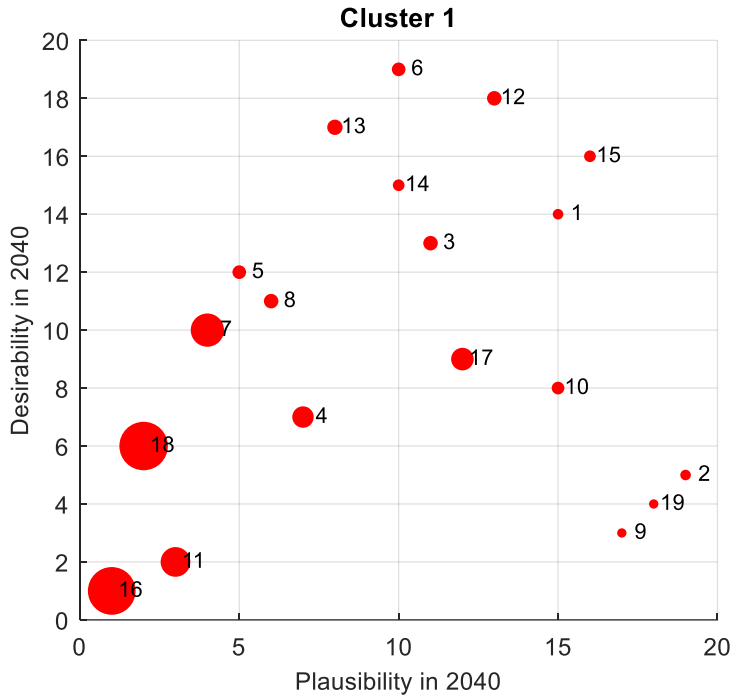

Cluster 2
Impact (on the online space in 2040)
Probability of occurrencein 2040
0 1 2 3 4 5 6 7 8 9 10
6
8
1
2
5
4
7
9
3

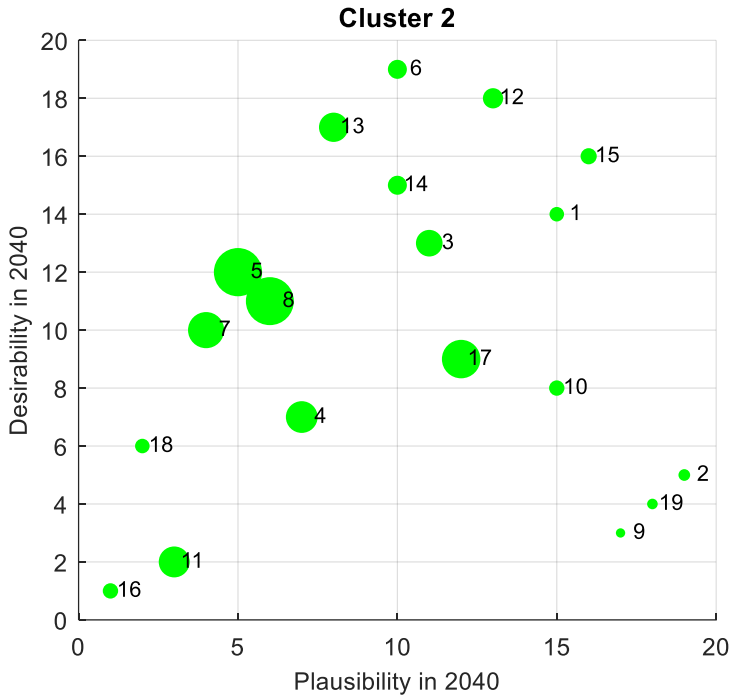

Cluster 3
Impact (on the online space in 2040)
Probability of occurrencein 2040
0 1 2 3 4 5 6 7 8 9 10
6
8
1
2
5
4
7
9
3

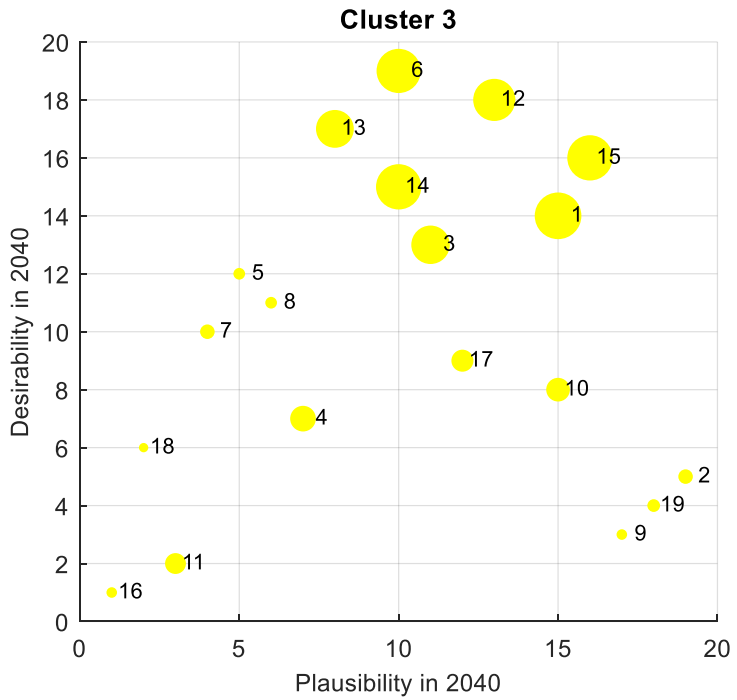

Cluster 4
Impact (on the online space in 2040)
Probability of occurrencein 2040
0 1 2 3 4 5 6 7 8 9 10
6
8
1
2
5
4
7
9
3

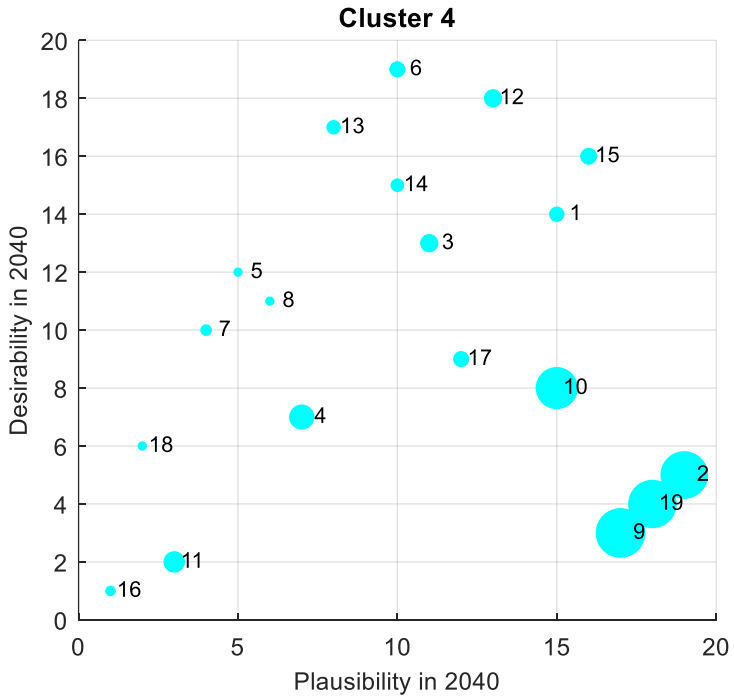
Fuzzy clustering reulsts: **Key factors of social change**

Clus 1	Clus 2	Clus 3	Clus 4	
0.02	0.06	0.85	0.07	1. New technologies in the future (not present today) will help in countering the phenomenon.
0.02	0.03	0.06	0.89	2. New technologies in the future will provide new opportunities for disseminating misogynistic hate messages.
0.06	0.26	0.57	0.11	3. Legislation on freedom of expression will offer new possibilities to "block" haters.
0.16	0.38	0.24	0.23	4. Legislation will offer more freedom of expression (therefore, it will ease the haters).
0.05	0.91	0.03	0.01	5. Psychological harm due to mis. h.s. will be recognized as a disease by public health and insurance comp.
0.05	0.12	0.76	0.08	6. Parents will devote themselves to educating their children on the phenomenon.
0.42	0.50	0.06	0.03	7. Parents will control (manually or digitally) everything their children post online.
0.06	0.90	0.03	0.01	8. Online misogynistic hate speech will drastically worsen the social relationships of those who make them.
0.01	0.01	0.02	0.97	9. Online misogynistic hate speech will drastically worsen the social relationships of those who suffer from them.
0.04	0.07	0.20	0.69	10. 90% of communication among young people will be “virtual”.
0.33	0.36	0.15	0.16	11. 90% of communication between parents and children will be “virtual”.
0.06	0.14	0.69	0.11	12. The school will play a crucial role in countering and preventing the phenomenon.
0.07	0.32	0.55	0.06	13. The mass media will play a crucial role in countering and preventing the phenomenon.
0.03	0.12	0.80	0.05	14. Public services (social, health, etc.) will assist victims of misogynistic hate speech.
0.03	0.08	0.80	0.09	15. Voluntary associations (parishes, non-profit organizations, etc.) will assist victims of misogynistic hate speech.
0.89	0.07	0.02	0.02	16. There will be more hate speech against men than against women.
0.18	0.57	0.17	0.08	17. There will be private companies getting paid to monitor and report haters.
0.92	0.06	0.01	0.01	18. The national government will impose a tax to support the cost of managing these phenomena.
0.01	0.02	0.04	0.92	19. Misogynistic hate speech spread on social media will be used to drive citizens' political choices.

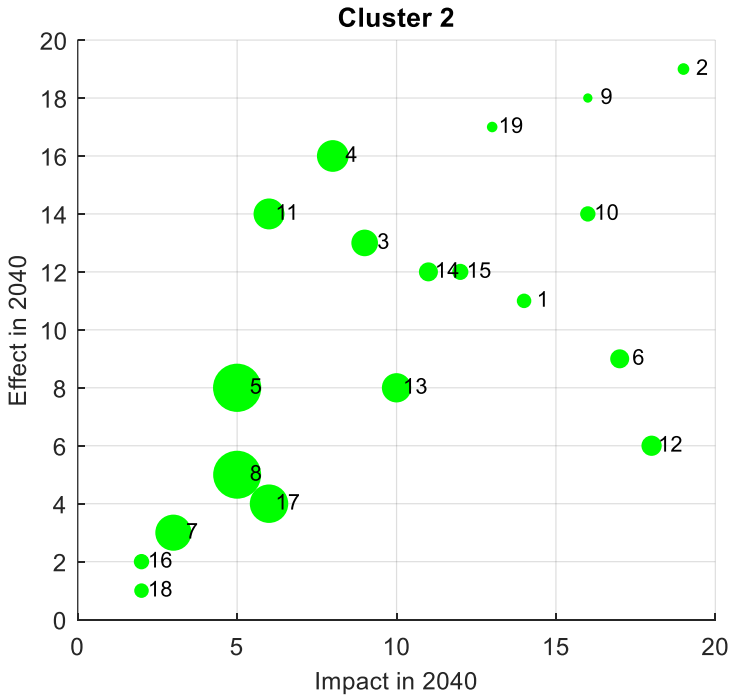

Cluster 1
Desirability in 2040
Plausibility in 2040
0
2
4
6
8
10
12
14
16
18
20
6
12
13
15
14
1
3
5
8
7
17
10
4
18
2
19
9
11
16

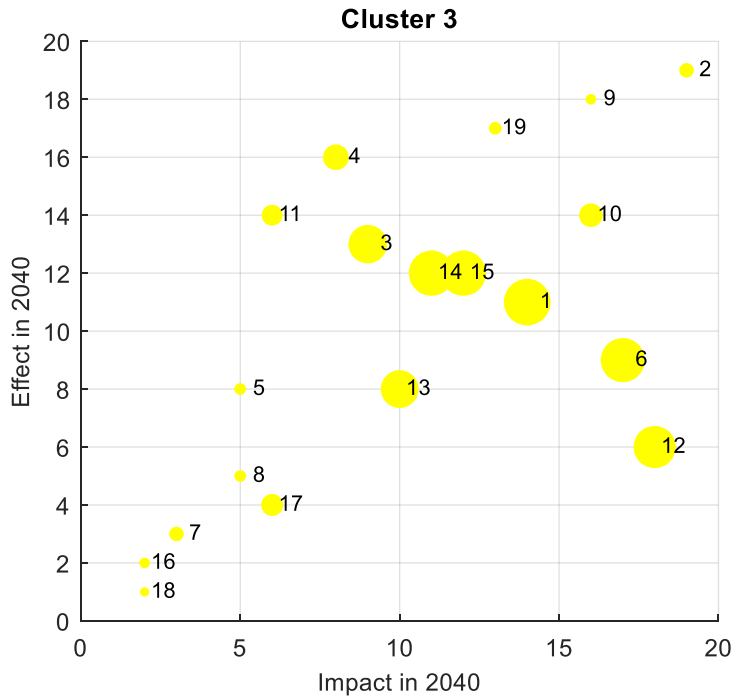

Cluster 2
Desirability in 2040
Plausibility in 2040
0
2
4
6
8
10
12
14
16
18
20
6
12
13
15
14
1
3
5
8
7
17
10
4
18
2
19
9
11
16

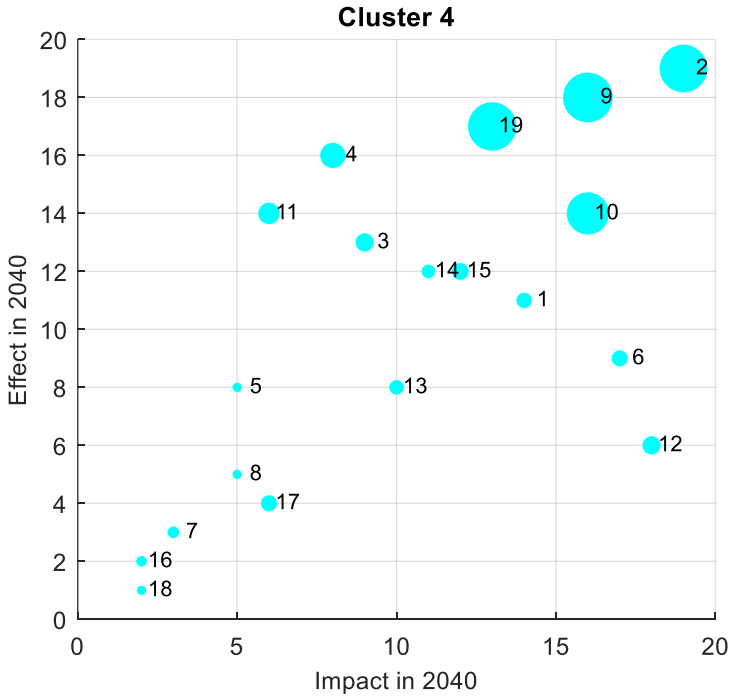

Cluster 3
Desirability in 2040
Plausibility in 2040
0
2
4
6
8
10
12
14
16
18
20
6
12
13
15
14
1
3
5
8
7
17
10
4
18
2
19
9
11
16


Cluster 4
Desirability in 2040
Plausibility in 2040
0
2
4
6
8
10
12
14
16
18
20
6
12
13
15
14
1
3
5
8
7
17
10
4
18
2
19
9
11
16

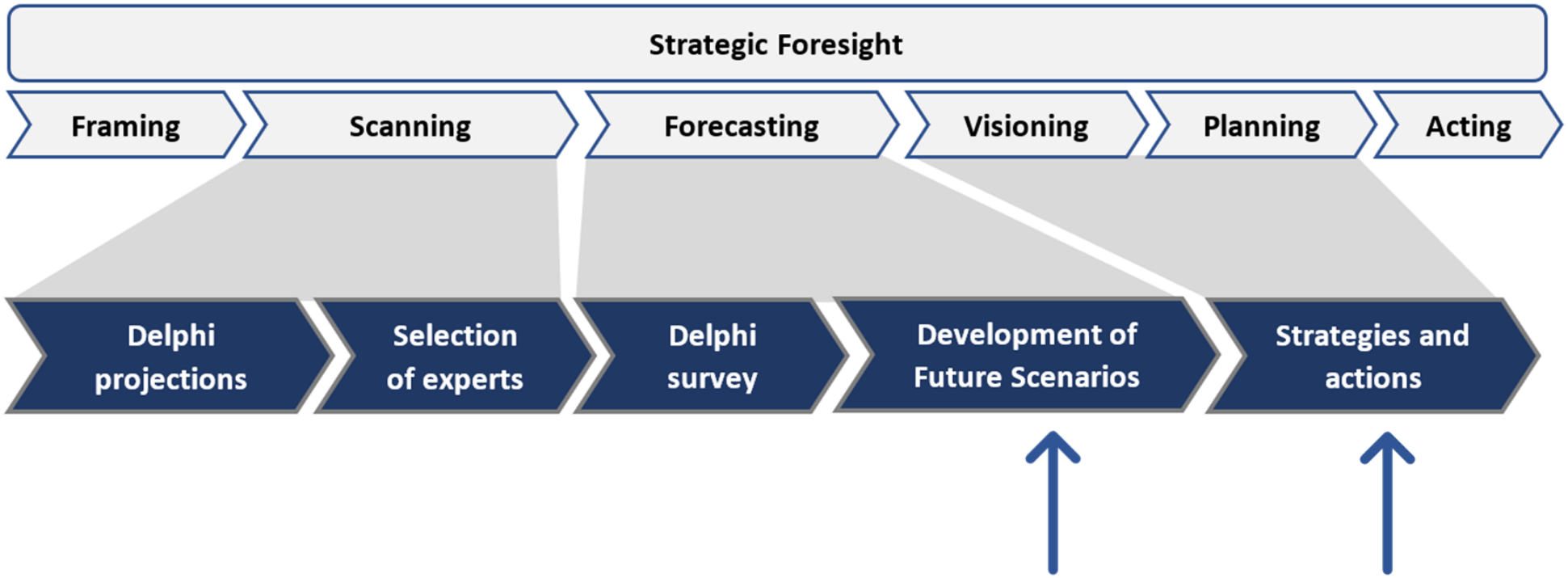
Cluster 1
Effect in 2040
Impact in 2040
0
2
4
6
8
10
12
14
16
18
20
2
9
19
4
11
10
3
14
15
1
6
5
13
12
8
17
7
16
18


Cluster 2
Effect in 2040
Impact in 2040
0
2
4
6
8
10
12
14
16
18
20
2
9
19
4
11
10
3
14
15
1
6
5
13
12
8
17
7
16
18


Cluster 3
Effect in 2040
Impact in 2040
0
2
4
6
8
10
12
14
16
18
20
2
9
19
4
11
10
3
14
15
1
6
5
13
12
8
17
7
16
18


Cluster 4
Effect in 2040
Impact in 2040
0
2
4
6
8
10
12
14
16
18
20
2
9
19
4
11
10
3
14
15
1
6
5
13
12
8
17
7
16
18

Future Developments



- Development of future scenarios
- Expert checks for plausibility and consistency

- List of policy actions
- Ranking and Evaluation of the policies actions

Thank you for your attention

References

- Di Zio S., Bolzan M., Marozzi M. (2021), Classification of Delphi outputs through robust ranking and fuzzy clustering for Delphi-based scenarios, Technological Forecast. & Soc. Ch., 173, 121140.
- Di Zio S., Tontodimamma A., del Gobbo E., Fontanella L.,(2023) Exploring the research dynamics of futures studies: An analysis of six top journals, Futures, Volume 153, 2023, 103232, ISSN 0016-3287, https://doi.org/10.1016/j.futures.2023.103232.
- Gagliardone, I.; Gal, D.; Alves, T.; Martinez, G. (2015): Countering Online Hate Speech. Paris: UNESCO Publishing.
- Marozzi M., Di Zio S., Bolzan M. (2022) Robust Weighted Aggregation of Expert Opinions in Futures Studies, Annals of Operations Research, DOI: 10.1007/s10479-022-04990-z.
- Richardson-Self, L. (2018): Woman-Hating: On Misogyny, Sexism, and Hate Speech. Hypatia, 33(2), 256-272.
- Siegel, A. A. (2020): Online hate speech. In J. Tucker & N. Persily (Eds.), Social Media and Democracy: The State of the Field. Cambridge: Cambridge University Press.
- Tontodimamma, A.; Nissi, E.; Sarra, A.; Fontanella, L. (2021): Thirty years of research into hate speech: topics of interest and their evolution. Scientometrics 126, 157–179.